HHS Public Access

Author manuscript

Med Image Anal. Author manuscript; available in PMC 2016 October 01.

Published in final edited form as:

Med Image Anal. 2015 October ; 25(1): 2–10. doi:10.1016/j.media.2015.04.004.

Shape Analysis Based on Depth-Ordering

Yi Hong[a], **Yi Gao**[c], **Marc Niethammer**[a,b], and **Sylvain Bouix**[d]

[a]Department of Computer Science, University of North Carolina (UNC) at Chapel Hill, NC, USA

[b]Biomedical Research Imaging Center, UNC-Chapel Hill, NC, USA

[c]Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, USA

[d]Psychiatry Neuroimaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Abstract

In this paper we propose a new method for shape analysis based on the ordering of shapes using band-depth. We use this band-depth to non-parametrically define a global depth for a shape with respect to a reference population, typically consisting of normal control subjects. This allows us to globally quantify differences with respect to "normality". Using the depth-ordering of shapes also allows the detection of localized shape differences by using α-central values of shapes. We propose permutation tests to statistically assess global and local shape differences. We further determine the directionality of shape differences (local inflation versus deflation). The method is evaluated on a synthetically generated striatum dataset, and applied to detect shape differences in the hippocampus between subjects with first-episode schizophrenia and normal controls.

Keywords

shape analysis; depth-ordering of shape; global analysis; local analysis

1. Introduction

Analyzing and comparing three-dimensional brain structures or objects in general can be as simple as comparing volumes. While informative, such a global measure cannot fully describe changes between objects. Shape analysis approaches have been proposed to assess object properties beyond global volume and to characterize shape variations across subjects and between subject populations (see Nitzken et al. (2014) for a brief survey). For instance, many shape analysis methods are based on the classical point distribution model (PDM) (Cootes and Taylor, 2004), which captures shape variations by computing a mean shape and the major modes around the mean of corresponding points in a set of shapes as illustrated in Fig. 1(a). The PDM assumes a Gaussian distribution of the points around the mean shape.

In contrast, shape characterizations built on concepts of order statistics have been explored recently (Whitaker et al., 2013; Hong et al., 2013, 2014a). These methods utilize depth-ordering of shapes to generalize order statistics, for example, the median and the inter-quartile range (IQR), to shapes, effectively obtaining the equivalent of a box-plot for shapes. Fig. 1(b) illustrates the depth-ordering model (DOM). Using shape-descriptions based on depth-ordering makes it possible to perform shape analysis without making strong distributional assumptions. While current approaches focus on using DOM to analyze shape variations within one population, this paper addresses the challenge of differentiating subject populations, for example subjects with a disorder versus normal controls, based on the depth-ordering model.

Existing methods for population-based shape analysis can be roughly subdivided into two categories: methods for global analysis and methods for local analysis. Global analysis methods are designed to detect whether population shape differences exist (Loncaric, 1998; Reuter et al., 2009; Wachinger et al., 2014), but they generally cannot locate or characterize these shape differences. This limits interpretability of results and consequentially insights into the underlying biological processes. The main attraction of such methods is that they often avoid establishing dense correspondences between shapes through registration. In contrast, local analysis methods require some form of point-to-point correspondence between shapes to allow precise local shape analysis. Establishing these correspondences is highly non-trivial and arguably one of the main sources of inaccuracy, because any misregistration may create artifacts with respect to the final shape analysis results. Nevertheless, a variety of methods for local shape analysis have been proposed and successfully used (Miller, 2004; Davies et al., 2008; Cates et al., 2008; Chung et al., 2008; Hosseinbor et al., 2014). Shape representations that fit an a-priori model to the data have also been used successfully, although they still need to establish some form of one-to-one correspondence via model fitting (Styner et al., 2004; Yushkevich and Zhang, 2013).

In this work, we explore an alternative method that allows for both global and local shape analyses, and only needs very limited (e.g., rigid or affine) spatial alignment of shapes. This is achieved by using the notion of band-depth (López-Pintado and Romo, 2009; Hong et al., 2013), which provides a notion of centrality of a shape with respect to a reference dataset. The deeper the shape the more similar it is to all other shapes in the dataset. We use this notion to detect small morphometric differences between two populations of neuroanatomical structures. Fig. 2 shows the general principle of the approach. First, shapes are depth-ordered with respect to a reference dataset. We then define a global statistical test directly on the shape depth, as well as a local statistical test to assess local shape differences. Our method is free from strong distributional assumptions by using principles from non-parametric order-statistics.

Our main contributions in this paper are:

1. We propose using depth-ordering on shapes for statistical shape analysis.
2. We develop an algorithm for the fast computation of band-depth for shapes represented by *binary* indicator functions.

3. We define statistical tests to detect potential global and local differences between shape populations without an explicit computation of dense correspondences.
4. We provide the directionality of shape differences to augment local measurements.

The method described in this paper is an extension of preliminary ideas presented in a recent conference paper (Hong et al., 2014b). This paper provides further details of our algorithms including the permutation tests for both global and local analyses, presents an additional technique for providing directionality of shape differences, and includes more comprehensive experiments on both synthetic and real datasets.

The paper is organized as follows: Section 2 describes how to order shapes by depth and how to compute such a depth-ordering fast. Section 3 proposes statistical approaches using depth-ordering for shape analysis. Section 4 discusses how to augment local analysis with directionality of shape differences. Section 5 presents experimental results on synthetic and real datasets. Section 6 concludes the paper with a summary.

2. Depth-ordering of shapes

Generalizing concepts from order-statistics to shape analysis faces the challenge that there is no canonical ordering of shapes. To define such an ordering we make use of the concept of band-depth and ordering of functions as developed in the statistics literature (López-Pintado and Romo, 2009) and extend it to shapes. Sun and Genton (2011) first proposed functional boxplots to order functions using the band-depth concept in López-Pintado and Romo (2009). The intuition of ordering functions based on band-depth is that the deeper a function is buried within a dataset the more central it is. The deepest function corresponds to the within-sample median function. Once defined, this ordering can be used to generalize traditional order statistics, such as the median or the inter-quartile range, to functions. For example, Whitaker et al. (2013) adapted traditional boxplots to *functional contour boxplots* for quantifying uncertainty in fluid simulations and for the visualization of ensemble data. What makes band-depth attractive for *shape*-ordering is that shapes can be represented by *indicator functions*, i.e., by binary functions that are 1 inside and 0 outside of a shape (Hong et al., 2013, 2014a).

Band-depth for binary shape representations relates to the amount of overlap between shapes, because the minimum and maximum operators on indicator functions correspond to the set intersections and unions respectively. Therefore, band-depth is a *natural and intuitive* choice to order a shape population for an indicator-function-based shape representation. In our current work, we use this binary function representation to compare shape populations. To improve the computational efficiency of our model, we propose a novel fast algorithm to compute the band-depth of shapes represented by binary maps. Most importantly we demonstrate how band-depth can be used to provide both global and local statistical tests to differentiate between shape populations.

Given a set of n shapes represented as 3D binary volumes, $\{Y_1, Y_2, \ldots, Y_n\}$, with dimension of (s_x, s_y, s_z), we vectorize them into binary vectors $y_i \in \{0, 1\}^p$ $(1 \leq i \leq n)$, where $p = s_x \times s_y \times s_z$. The band-depth for each shape y is defined as follows:

$$BD_{n,j}(\mathrm{y}) = \sum_{j=2}^{J} BD_n^{(j)}(\mathrm{y}) \quad where \quad BD_n^{(j)}(\mathrm{y}) = \frac{1}{C} \sum_{1 \le i_1 < i_2 < \cdots < i_j \le n} I\{G(\mathrm{y}) \subseteq B(\mathrm{y}_{i_1}, \cdots, \mathrm{y}_{i_1})\} \ . \tag{1}$$

Here, $2 \le j \le J$, and $J \in [2, n]$ is the number of observations used to define the band. C is a normalization constant equal to the number of admissible permutations, i.e., $\binom{n}{j}$. $G(\mathrm{y})$ is the graph of the function, $G(\mathrm{y}) = \{\mathbf{x}, \mathrm{y}(\mathbf{x}) : \mathbf{x} \in \mathbb{I}\}$, and $\mathbb{I}$ denotes the index range of the binary vector y. B is delimited by the observations given as its arguments. That is, $B(\mathrm{y}_{i_1}, \ldots, \mathrm{y}_{i_j}) = \{(\mathbf{x}, \mathrm{y}(\mathbf{x})) : \mathbf{x} \in \mathbb{I}, min_{r=i_1,\ldots,i_j} \mathrm{y}_r(\mathbf{x}) \le \mathrm{y}(\mathbf{x}) \le max_{r=i_1,\ldots,i_j} \mathrm{y}_r(\mathbf{x})\}$. $I\{\cdot\}$ denotes the indicator function, which evaluates to 1 if the graph of the function is within the band, or to 0, otherwise.

Intuitively, the more central a particular vector y is in the data population the more likely it will be “surrounded” by other vectors; and hence it results in many evaluations of the indicator function to 1 and therefore a large value for the band-depth. Conversely, a vector at the fringes of the data population will rarely be surrounded by other vectors and therefore will be assigned a comparatively low value for its band-depth.

Using the indicator function to define depth has limitations. For example many ties in depth may exist among shapes within the population. To address this issue, the band depth can be modified (Sun and Genton, 2011) to

$$MBD_{n,j}(\mathrm{y}) = \sum_{j=2}^{J} MBD_n^{(j)}(\mathrm{y}) \quad where \quad MBD_n^{(j)}(\mathrm{y}) = \frac{1}{C} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} \lambda_p\{A(\mathrm{y}; \mathrm{y}_{i_1,\ldots,\mathrm{y}_{i_j}})\} \ , \tag{2}$$

where $A_j(\mathrm{y}) \equiv A(\mathrm{y}; \mathrm{y}_{i_1}, \ldots \mathrm{y}_{i_j})$ and $A_j(\mathrm{y}) \equiv \{\mathbf{x} \in \mathbb{I} : min_{r=i_1,\ldots,i_j} \mathrm{y}_r(\mathbf{x}) \le \mathrm{y}(\mathbf{x}) \le max_{r=i_1,\ldots,i_j} \mathrm{y}_r(\mathbf{x})\}$, $\lambda_p(\mathrm{y}) = \lambda(A_j(\mathrm{y}))/\lambda(\mathbb{I})$, λ is the Lebesgue measure on $\mathbb{R}^p$ and p is the observation’s dimension, i.e., the number of voxels in a binary represented shape.

In other words, for shapes represented as discrete binary vectors, computing the modified band depth (MBD) of a shape is equivalent to calculating the depth of each one of its element and reporting an average of the proportion of the shape contained within bands. This is different from the original band-depth (BD, see Equation (1)), which computes an average of being *fully* contained within bands. Because measuring the proportion of being contained within a band provides more information about the relationship of a shape with a band, MBD has fewer ties in the resulting depth values than BD. Hence, in our work we used this modified measure in all the experiments.

Albeit its conceptual simplicity, one of the main limitations of band-depth is its computational complexity. Take $J = 2$ for example, that is, the band is defined by two observations[1]. Because each observation needs to be compared with the band formed by two out of n observations, the original algorithm (Sun and Genton, 2011) has $O(pn^3)$ complexity for computing band-depth for all observations, as illustrated in Fig. 3(a). This makes this algorithm unsuitable for large numbers of functions. To allow for faster computations recently a fast method to compute band-depth has been proposed (Sun et al., 2012). This method is based on computing local curve ranks and requires sorting of values from all curves at the same position. The algorithm uses a rank list to compute the frequency with which one point of a curve is contained by other curves at the same location. However, the proposed algorithm is ill-suited for binary shape representations as it does not consider ranking ties during sorting, which frequently appear when sorting binary values. Fig. 3(b) illustrates this ranking approach.

We observed that for binary representations sorting can be avoided, as at any location only two values are possible, that is, no sorting, but simple addition and subtraction, is enough to measure whether a point of a shape is contained by the band formed by shapes of a population at that location. The computation of MBD can then be accomplished efficiently for $J = 2$ using our algorithm as follows:

S0) Given n binary volumes, $\{\mathrm{Y}_i\}_{i=1}^{n}$, vectorize them: $\mathrm{y}_i \in \{0, 1\}^p$, p is the number of voxels in a binary volume.

S1) At each location of a volume, k ($1 \le k \le p$) for a given value $v(k) \in \{0, 1\}$, we count the number of functions that have a value larger (n_a), smaller (n_b) or equal (n_t) to $v(k)$:

- **a.** if $v(k) = 0$, then $n_a = \sum_{i=1}^{n} \mathrm{y}_i(k)$, $n_b = 0$, and $n_t = n - n_a - 1$;
- **b.** if $v(k) = 1$, then $n_a = 0$, $n_b = \sum_{i=1}^{n} (1 - \mathrm{y}_i(k))$ and $n_t = n - n_b - 1$.

This procedure allows us to evaluate how often a voxel of a volume is larger (or smaller) than voxels of other volumes at the same location.

S2) Based on the numbers n_a and n_b defined in step S1, we can calculate the number of bands containing $v(k)$, for example, from y_i. First, let B be a band such that y_i is not one of its observations, and its bounds at k are denoted as b_{min} and b_{max}. For $J = 2$, we have the following possible scenarios, i.e., the combinations of possible B that contain $v(k)$:

- **a.** $b_{min} < v(k) < b_{max}$ ($n_a n_b$ such combinations).
- **b.** $b_{min} = v(k)$ and $v(k) < b_{max}$ ($n_a n_t$ such combinations).
- **c.** $b_{min} < v(k)$ and $v(k) = b_{max}$ ($n_b n_t$ such combinations).
- **d.** $b_{min} = v(k)$ and $v(k) = b_{max}$ ($n_t(n_t - 1)/2$ such combinations).

[1]Due to computational complexity this is typically how band-depth is evaluated, because choosing $J > 2$ results in even higher computational costs.

In addition, we have to account for bands that have y_i as one of its observations. There are $n - 1$ such bands. By collecting all combinations the total number of bands ($J = 2$) containing the value $v(k)$ is:

$$C_k(v(k))=n_a n_b+n_a n_t+n_b n_t+n_t(n_t-1)/2+(n-1).$$

For binary functions n_a and n_b cannot simultaneously be different from zero, which simplifies the expression to

$$C_k(v(k))=(n_a+n_b)n_t+n_t(n_t-1)/2+n-1.$$

The above equation provides us with the number of times a voxel of a volume is contained by voxels of any two volumes at the same location.

S3) The modified band-depth for a shape y_i is then

$$MBD(y_i)=\frac{1}{p}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{k=1}^{p} C_k(y_i(k)),$$

where the notation $C_k(y_i(k))$ denotes computing C_k in step S2 based on the coefficients n_a, n_b, n_t given by the value of y_i at location k.

We can connect this special case scenario to the original MBD definition (Equation (2)) as follows. First, for bands made of two observations we have $J = 2$ and thus $C=\begin{pmatrix} n \\ 2 \end{pmatrix}$. Second, for volumes of size p, we have $\lambda(\mathbb{I})=p$. Finally, for binary volumes $\sum_{1\le i_1<i_2<i_j\le n}\lambda(A_j(y_i))$ is equivalent to $\sum_{k=1}^{p} C_k(y_i(k))$.

This algorithm computes MBD of all shapes in a population at the same time, not sequentially. Compared to the original band-depth algorithm our computational complexity is reduced from $O(pn^3)$ to $O(pn)$, as shown in Fig. 3. This makes it possible to compute band-depth for large populations and large multi-dimensional shapes (we focus on 2D surfaces in 3D space in this paper).

Note that, when choosing two shapes from a population to define a band, we do not exclude the one that is being evaluated from the the admissible permutations of the population. Instead, all the shapes in the population will be considered to define the bands. So, in the algorithm the number of observations is n, i.e., the number of shapes in a population. One can also exclude the current shape and compute its depth value with respect to a population of the remaining shapes. This will not change the ranking because the current shape is always contained within the band formed by shapes including itself, resulting in the same constant difference in the depth value of each shape, compared to using n shapes as a population in the algorithm.

3. Statistics using depth-ordering

Band-depth computed in Section 2 measures the relationship between a shape and a reference population. A higher value indicates the shape is closer to the median of the reference population, and a lower one indicates the shape is a potential outlier with respect to the reference population. Based on this property of band-depth we can perform global shape analysis as described in Section 3.1 as well as local shape analysis as described in Section 3.2. For all these analyses we assume that shapes have been pre-aligned as appropriate. Typically, this will either involve a rigid, similarity or affine alignment of shapes to a given template or to a template obtained by some form of unbiased atlas-building method. The choice of transform will depend on the objective of a given study. For example, if size differences should be included, rigid alignment would be appropriate.

The key ingredient to performing statistics using depth-ordering is to compute depth-ordering with respect to a reference population of shapes that are used as a non-parametric model of shapes. In particular, while band-depth would typically be computed for each shape of a population with respect to all the shapes, in our definition of statistical tests we will make use of a reference population to which other shapes are compared, one at a time. Intuitively, one establishes a reference population (for example a population of normal control subjects) and then for another given shape not part of this population, one tests how deep this shape would be buried with respect to the reference. Note that if a new subject is added into the reference population, the band-depth of each shape should be recomputed with respect to the new reference population.

3.1. Global shape analysis

The goal of global shape analysis is to establish if there are shape differences between two different groups. To this end we assign a scalar description of shape, here a measure of depth, to each shape. Given a dataset $\{R_i\}$ containing training population of shapes, we compute the band-depth for a given datum D, from a set of input test shapes $\{D_j\}$. We first compute the band-depth for all the data in a population of $\{R_i\} \cup D$ using the algorithm described in Section 2, and then assign the resulting band-depth of D to the datum D, denoted as $BD(D; \{Ri\})$, as illustrated in Fig. 4. A larger depth value indicates the test shape being closer to the training set. Typically, we choose a subgroup of control subjects as the training population, so in general, the control test set would have larger depths than the disease test set.

In the proposed method the training population forms a “yard-stick” by which to judge data-depth. This is substantially different from directly computing the band-depths for the full dataset $\{Dj\}$, which would be problematic. Most easily this effect is illustrated by considering scalar-valued data of two groups with the same variance, but strongly different means. The most central element (the median) in such a case would be half-way between the means. This median element will have the largest band-depth. Band-depth will decrease moving away from this median element towards smaller *and* larger elements. Hence, the groups would become indistinguishable even though they are clearly different. Equally problematic would be to compute band-depth *separately* for both groups in which case they cannot be meaningfully compared directly as the band-depth values become relative to each

group. To illustrate this effect Fig. 5 shows the results of the proposed approach using a reference population and the approach using separate band-depth computations for the groups of the synthetic striatum data described in detail in Section 5.1. Our proposed approach using a reference group can clearly differentiate the populations whereas a computation of the band-depth without a reference group is not successful.

3.1.1. Permutation test—To measure if global shape differences are statistically significant between control and disease groups, we use a permutation test on the mean depth of the control group versus that of the disease group. The null hypothesis of the permutation test is that both the control and the disease groups have the same mean depth with respect to the training group. Specifically, since the depth of each test shape has been computed with respect to the training group, we first compute the mean depth difference between the control test group and the disease test group with no permutation. Then, in each permutation we exchange the subjects in the two test groups and compute the mean depth difference between the two permuted test groups[2]. We count the number of times that the mean depth difference is larger than the one computed without permutation, and the proportion of larger values to the total number of values is the associated *p*-value. We test for statistical significance at a level of 0.05, i.e., we declare statistical significance for a *p*-value ≤ 0.05.

3.2. Local shape analysis

For the local shape analysis the goal is not only to establish whether there are shape differences, but also where these shape differences occur. To better understand the local analysis, we first introduce its key ingredient, the α-central region as shown in Fig. 6, which allows us to define a local shape analysis approach using depth-ordering. Specifically, based on the ordering of shapes, we can not only compute the median shape, the most central shape of a population, but also α-central regions ($0 < \alpha \leq 1$), similar to for example the inter-quartile-range (IQR) for scalar-valued cases. The band of the α-central region is delimited by the α proportion of the deepest shapes in the ranking list, which is defined as

$$CR_\alpha = \{B(\mathrm{y}_{[1]}, \mathrm{y}_{[2]}, \ldots, \mathrm{y}_{[q]}), q = \lceil \alpha n \rceil\}, \quad (3)$$

where $\mathrm{y}_{[1]}, \ldots, \mathrm{y}_{[q]}$ ($q \geq 1$) denote the ordered shapes, from the deepest one to the qth-deepest one, $B(\cdot)$ is the band as defined in Section 2. The grey band shown in Fig. 6 illustrates a central region with α equal to 0.6, because the three deepest shapes, out of a total of five shapes of a population, are used to build this central region.

The local analysis is based on the α-central regions of a reference population to detect the location of shape differences. For the shapes in the reference population, starting from the deepest shape (its median shape) we gradually add shapes in order of decreasing band-depth. Each set of shapes defines a specific α-central region and a particular location is assigned the α value of the first α-central region that covers it. The resulting map describes the "centrality" of a shape population at each point in the domain. The test shape can be overlaid

[2]Since the permutation is performed between test groups and there is no change in the training group, the depth value of each test subject will keep unchanged during this process. Hence, we do not need to recompute their band-depth values.

on this centrality map and the corresponding α-values will be recorded on its surface, thus providing a local measure of shape abnormality.

Fig. 7 illustrates this concept for a population of two-dimensional shapes. Given the reference shapes, the blue shapes shown in Fig. 7(a), we build a centrality map based on the band-depth of each shape. As shown in Fig. 7(b), the deepest shape has the lowest α-value (light yellow) and the most outlying shape has the highest α-value (dark red). A local measure of "belonging" to the population can then be computed for a test shape by tracing the α-central region it traverses, see Fig. 7(c). Note that some regions of the shape may not be covered by the reference shape population, for example the top right part of the test shape. To assign an α value in such cases, we use a dilation procedure starting from the boundary of the $\alpha = 1$ central region, evolving at a constant speed until all locations of the volume are covered. This results in α values greater than 1, shown as the regions colored with the darkest red in Fig. 7.

Overall, $\alpha > 0$, and a larger α-value indicates a local region with a higher potential to be different with respect to the reference group. Note that the centrality map has no notion of the directionality of shape differences. Instead, it captures both the inflated and deflated shape differences with respect to the reference group without distinguishing them, as the bump and the indented region shown in Fig. 7.

3.2.1. Permutation test—To measure whether a local region of a test group is significantly different from another test group, we design a permutation test using the α-values. Specifically, similar to the global analysis, we have three groups of subjects: a group of control subjects as the training set, another group of control subjects as the control test set, and a group of disease subjects as the disease test set. We first compute the band-depth for each shape from the two test groups with respect to the training set. That is, each test shape will be assigned a band-depth value, and its band-depth will be used for ordering in the permutation test. Then, the median of one test group is chosen as the template, and the other test group is left as the reference population. Our goal is to measure the shape differences of the template compared to the reference population, which is ordered based on the associated band-depths of its shapes when computing the α-values for the template.

We perform a permutation test with the null hypothesis being that the α-values recorded on the surface of the template shape are the same for the two test populations. In each permutation, we exchange the subjects in the two test groups to reconstruct a new reference group and reorder its shapes according to the band-depth associated with them. Note that we do not recompute these band-depths, but always keep the band-depth values that are associated to a particular shape. Then, new α-values for the template with respect to the new reference population are computed. At each position we count the number of the α-values that are larger than the one computed without permutation, resulting in *p*-values on the surface of the template. A *p*-value smaller than 0.05 (with 10000 permutations for example) indicates that the corresponding local region of the template is significantly different from that region of the reference population at a significance level of 0.05.

4. Directionality of shape differences

While the concept of band-depth allows us to measure how central a shape is with respect to a reference population, this measure is not signed and thus cannot represent whether a particular region of a shape is inflated or deflated (i.e., atrophied). In other words, band-depth ordering lacks *directionality*. For example, in Fig. 7 two different regions can be flagged as abnormal using the centrality map, but it is not possible to tell that one is ”thinner” and the other ”thicker” purely based on band-depth or the assigned α-values.

To measure the directionality of shape differences, we propose to use the signed distance transform. The signed distance map of the median shape of the reference population is computed. The test shape is then overlaid onto this signed distance map, so that the corresponding values can be recorded on the surface of the test shape, as illustrated in Fig. 8. As a result, a positive value indicates an enlarged region with respect to the reference median, and a negative value indicates an atrophied region.

5. Experimental results

We applied our method on both synthetic and real data. All shapes are pre-aligned using a rigid transformation, and represented as binary functions.

5.1. Synthetic data

Using synthetic data allows us to introduce a predefined shape change which we wish to recover using our proposed approach. We use the technique described in Gao and Bouix (2012) to generate large data sets of realistic shapes with known deformations. In short, a manifold learning technique is used to generate arbitrarily many shapes from a small training sample. A joint clustering algorithm is then applied to parcellate each shape’s surface into small regions which are consistently located across all shapes. Finally, a Log-Euclidean framework is used to introduce smooth, invertible and anatomically realistic deformations to one or multiple regions as defined by the clustering.

For this application, we generated 160 shapes based on 27 manually traced striata. We then modified 80 of them by thickening the putamen. A sample shape with deformation distance is shown in Fig. 9. Here, the medial side of the putamen was “pulled out”, and because of this large deformation the lateral side of the putamen was slightly deflated due to the diffeomorphic nature of the deformation. We evenly divided 80 normal controls into two groups. One group is used as the training set (NC-Train), and the other for testing (NC-Test). From the 80 abnormal subjects, we randomly picked 40 of them for testing.

5.1.1. Global analysis—To test for global group separability, we followed the strategy described in Section 3.1.1 to perform a permutation test (10000 permutations). When using the NC-Train as the training set to compute band-depth for both control and disease test sets, the resulting *p*-value is less than 1e-4, indicating the normal controls and the disease subjects are significantly different. On the other hand, as shown in Fig. 5, when pooling all shapes together to compute their band depths, no significant difference is detected. Since we know the control and disease groups of the synthetic data are significantly different according to

the ground truth, this experimental result indicates that the training population is essential for the discrimination of subject populations.

5.1.2. Local analysis—For the local analysis, we first need to choose the template and the reference population for the permutation test. Two possible choices are: (i) taking the median of the NC-Test as the template and the disease test group as the reference population, and (ii) using the median of the disease test group as the template and the NC-Test group as the reference population. We demonstrate our method using both strategies with experimental results shown in Fig. 10 and Fig. 11.

We first use the NC-Train group to compute the band-depth for all test shapes, and then estimate α-values of the template with respect to the reference population which is ordered using the computed band-depth. The α-values on the NC-Test median are shown in Fig. 10(a) and those of the disease median are shown in Fig. 11(a). Both results demonstrate that the introduced deformed (or different) regions on both medial and lateral sides of the template are detected as expected.

Fig. 10(b) and Fig. 11(b) show the raw *p*-values and the false-discovery-rate (FDR) corrected *p*-values, with 10000 permutations. Based on the ground truth shown in Fig. 9, the "pulled-out" region on the medial side of the putamen is significantly deformed, while the deflated region on the lateral side is slightly deformed, and some significantly deformed region may exist. We detect, regardless of which median shape is chosen as the template, the significantly deformed region. The main difference is that Fig. 10(b) shows more significantly different regions on the lateral side of the putamen than Fig. 11(b). Both strategies provide reasonable results compared to the ground truth (Fig. 9).

5.1.3. Directionality of shape differences—To test our strategy of measuring the directionality of shape differences, we first applied it on the median shapes of NC-Test and disease groups with respect to the median of NC-Train, as shown in Fig. 12. As expected, the median shapes from NC-Train and NC-Test have relatively small differences, c.f. Fig. 12(a). But, in Fig. 12(b) (the disease median with respect to the NC-Train median), we can see positive values on the medial side of the putamen, indicating that the disease median has an inflated region compared to the NC-Train median, and negative values on the lateral side of putamen, indicating that the disease median has a deflated region there. This is consistent with the introduced deformation in the synthetic data.

Since the directionality test among the control and the disease median shapes correctly reveals the directions of their shape differences, according to the introduced deformations, we further applied the directionality computation on the template with respect to the median shape of the reference population to augment the local analysis results based on depth-ordering. Fig. 10(c) shows the NC-Test median with respect to the disease median. The negative values on the medial side of the putamen indicate that this region of the NC-Test median (template) is deflated compared to the corresponding part of the disease median (from the reference population). And the positive values on the lateral side of the putamen indicate that this region of the NC-Test median is inflated compared to that part of the disease median. This is also consistent with the ground truth. By also considering the *p*-

values in Fig. 10(b), we can determine the directionality for those significantly different regions. Consistent results are shown in Fig. 11(c) for the disease median with respect to the NC-Test median.

5.2. Real data

Magnetic Resonance Images (MRI) of the brains of 123 subjects (including 102 males) diagnosed with first-episode schizophrenia and of 56 normal control subjects (including 37 males) were acquired on a 1.5-T scanner. Multi-site SPGR T1-weighted images (voxel dimensions 0.9375×0.9375×1.5 mm) were obtained. These MRI images were rigidly aligned to a standard coordinate space, from which hippocampus structures were segmented. Each binary segmentation was fit with a mesh model. This dataset was used in a previous shape analysis study (McClure et al., 2013). All the hippocampus shapes were pre-aligned using a rigid transformation. To get the binary representation of the hippocampus shapes, we used voxelizations with equal spacing in each direction. To measure the sensitivity with respect to spacing we tested using spacings of 0.3 and 0.4 mm. Results were similar. Hence, only results for a spacing of 0.4 mm are presented in what follows.

5.2.1. Global analysis—Similar to the synthetic data experiment, we divided the 56 normal control subjects into two groups, NC-Train and NC-Test, and considered all 123 subjects with first-episode schizophrenia as the disease group. Fig. 13 shows the global differences between NC-Test and the disease groups, for both left and right hippocampi. We also used a permutation test with 10000 permutations to determine whether these groups' mean depth values differed significantly. This resulted in *p*-values of 0.03 for the left hippocampus and 0.15 for the right hippocampus. This indicates based on the global depth-based analysis, that the left hippocampus is significantly different in disease and normal control populations at a significance level of 0.05.

5.2.2. Local analysis—For the local shape analysis, we used the NC-Train group as the training set to compute the depth for test subjects, and took the median of the disease group as the template and the NC-Test group as the reference population. Fig. 14 shows the local analysis results on both left and right hippocampi. We can see that our method captures some abnormal regions. However, based on the local *p*-values only relatively small regions of the median disease shape are significantly different from the normal controls. The *p*-values also show that the response is stronger on the left hippocampus, which is consistent with a previous study by McClure et al. (2013). The shown *p*-values for this experiment were not corrected using FDR. No significant regions remained after FDR-correction. Note that in our case we work directly on the voxel level of shapes thereby generating many local comparisons. Considering the analysis at a coarser spatial scale, e.g., averaged responses over regions as adopted in McClure et al. (2013), could potentially reveal shape differences which persist under FDR correction.

5.2.3. Directionality of shape differences—The right column in Fig. 14 shows the directionality of shape differences on the template (the disease median) with respect to the median of the reference group, NC-Test. We observe for both the left and right hippocampi of the disease median that few detected regions are locally deflated with respect to the

normal median, but instead most of the other detected regions are found to be locally inflated compared to the normal median.

5.3. Comparison with other methods

In Gao et al. (2014), three standard methods, SPHARM-PDM, Shapeworks, and Tensor Based Morphometry (TBM), were tested on the same synthetic dataset that we are using in the experiments. To quantitatively measure the significantly deformed region, they extracted the "region of deformation" (ROD) from ground truth data, and computed the ratio between the area of deformation and the area of the entire shape. The average of these ratios is referred to as the ground truth. For the striatum dataset used in our work the average ratio is 0.33. When measuring the detected significant region in the local analysis results, the ROD is defined as the region whose FDR-corrected *p*-value is less than or equal to 0.05. The area ratio reported in Gao et al. (2014) is 0.61 for SPHARM-PDM, 0 for ShapeWorks, and 0.17 for TBM.

The ROD cannot measure how well a shape analysis method localizes the differences, but it reveals whether the detected significant region has a reasonable size compared to the ground truth. Using the combination of ROD with the visual results allows one to assess both the location and extent of the detected deformation with respect to the ground truth. We follow this measurement and compute the area ratio of our FDR-corrected results on the synthetic data, resulting in 0.2 when using the median of NC-Test as the template (see Fig. 10), and 0.15 when using the median of the disease test group as the template (see Fig. 11). According to the ground truth of the area ratio for the synthetic data, our method has better performance than SPHARM-PDM and Shapeworks, and provides comparable, even slightly better, quantitative measures than TBM.

Compared to the qualitative results of the three methods presented in Gao et al. (2014), our method provides more accurate locations of the detected deformations, according to the introduced deformations. Furthermore, in contrast to these methods, our results reveal that most of the significantly different regions are located on the medial side of the putamen, instead of the lateral side, which is also consistent with the synthetic data.

In addition, our real data experiment is consistent with previous results on shape differences in the hippocampus for first-episode schizophrenia (McClure et al., 2013), specifically with strong differences for the left hippocampus.

6. Discussion and conclusion

In this paper we presented a shape analysis framework that can provide both global and local information, yet does not require complex processes to establish point-to-point correspondences. Instead we use the notion of band-depth of functions to order shapes according to how well they "fit in" a shape ensemble. This method allows for the definition of a median and α-central regions of a population, which can then be used to compare different population of shapes without strong distributional assumption.

Different from Hong et al. (2013), which focuses on augmenting a population atlas with statistical information using weighted band-depth, we proposed a fast algorithm to compute the band-depth of shapes represented by binary maps, and most importantly showed how band-depth can be used to provide both global and local statistical tests to differentiate between populations. In contrast to other deformation based tools for shape analysis, our approach is non-parametric and naturally captures how likely a shape belongs to a population. Although it does not provide physical measurements of displacement, these can be computed by deformation or a distance transform to the population median.

One limitation of our method is the reliance on a representative training set. Also, the size of the training set will affect the computed band-depth. Determining an appropriate size of the training set to obtain sufficient statistical power is left for future work.

In addition, as discussed our current method works on the pixel/voxel level. In future work we will explore a multi-scale approach to better adapt the analysis results to the spatial scale of the expected differences. To further increase the statistical power of our method, we could include more factors, e.g., age, gender, into the shape analysis, and explore their relationship with the estimated band-depths of shapes. This extension is not straightforward and therefore left to future work.

Acknowledgments

This work was supported by NIH grants R01-MH082918, P41EB002025, and R01-HL105241. We would like to thank James J. Levitt and Martin Styner for providing the datasets used in this paper.

References

Cates J, Fletcher PT, Styner M, Hazlett HC, Whitaker R. Particle-based shape analysis of multi-object complexes. Med Image Comput Comput Assist Interv. 2008; 11:477–485. [PubMed: 18979781]

Chung MK, Dalton KM, Davidson RJ. Tensor-based cortical surface morphometry via weighted spherical harmonic representation. IEEE Trans Med Imaging. 2008; 27:1143–1151. [PubMed: 18672431]

Cootes, T.; Taylor, C. Statistical models of appearance for computer vision. 2004.

Davies, R.; Twining, C.; Taylor, C. Statistical Models of Shape: Optimisation and Evaluation. Springer; London: 2008.

Gao, Y.; Bouix, S. Synthesis of realistic subcortical anatomy with known surface deformations. In: Levine, JA.; Paulsen, RR.; Zhang, Y., editors. Mesh Processing in Medical Image Analysis. Springer; Heidelberg: 2012. p. 80-88.

Gao Y, RiklinRaviv T, Bouix S. Shape analysis, a field in need of careful validation. Human brain mapping. 2014

Hong, Y.; Davis, B.; Marron, JS.; Kwitt, R.; Niethammer, M. Weighted functional boxplot with application to statistical atlas construction. In: Mori, K.; Sakuma, I.; Sato, Y.; Barillot, C.; Navab, N., editors. MICCAI 2013, Part III LNCS. Springer; Heidelberg: 2013. p. 584-591.

Hong Y, Davis B, Marron JS, Kwitt R, Singh N, Kimbell JS, Pitkin E, Superfine R, Davis SD, Zdanski CJ, Niethammer M. Statistical atlas construction via weighted functional boxplots. Medical image analysis. 2014a; 18:684–698. [PubMed: 24747271]

Hong Y, Gao Y, Niethammer M, Bouix S. Depth-based shape-analysis. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014). 2014b:17–24.

Hosseinbor AP, Kim WH, Adluru N, Acharya A, Vorperian HK, Chung MK. The 4D hyperspherical diffusion wavelet: A new method for the detection of localized anatomical variation. Med Image Comput Comput Assist Interv. 2014; 17:65–72. [PubMed: 25320783]

Loncaric S. A survey of shape analysis techniques. Pattern Recognition. 1998; 31:983–1001.

López-Pintado S, Romo J. On the concept of depth for functional data. Journal of the American Statistical Association. 2009; 104:718–734.

McClure RK, Styner M, Maltbie E, Lieberman JA, Gouttard S, Gerig G, Shi X, Zhu H. Localized differences in caudate and hippocampal shape are associated with schizophrenia but not antipsychotic type. Psychiatry Research: Neuroimaging. 2013; 211:1–10. [PubMed: 23142194]

Miller MI. Computational anatomy: Shape, growth, and atrophy comparison via diffeomorphisms. NeuroImage. 2004; 23:S19–S33. [PubMed: 15501089]

Nitzken MJ, Casanova MF, Gimelfarb G, Inanc T, Zurada JM, El-Baz A. Shape analysis of the human brain: a brief survey. IEEE J Biomed Health Inform. 2014; 18:1337–1354. [PubMed: 25014938]

Reuter M, Wolter FE, Shenton M, Niethammer M. Laplace–beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. Computer-Aided Design. 2009; 41:739–755. [PubMed: 20161035]

Styner M, Lieberman JA, Pantazis D, Gerig G. Boundary and medial shape analysis of the hippocampus in schizophrenia. Medical Image Analysis. 2004; 8:197–203. [PubMed: 15450215]

Sun Y, Genton M. Functional boxplots. Journal of Computational and Graphical Statistics. 2011; 20:316–334.

Sun Y, Genton MG, Nychka DW. Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? Stat. 2012; 1:68–74.

Wachinger C, Golland P, Reuter M. Brainprint: Identifying subjects by their brain. Med Image Comput Comput Assist Interv. 2014; 17:41–8. [PubMed: 25320780]

Whitaker RT, Mirzargar M, Kirby RM. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. IEEE Transactions on Visualization and Computer Graphics. 2013; 19:2713–2722. [PubMed: 24051838]

Yushkevich PA, Zhang HG. Deformable modeling using a 3D boundary representation with quadratic constraints on the branching structure of the Blum skeleton. Inf Process Med Imaging. 2013; 23:280–291. [PubMed: 24683976]

Highlights

- We propose using depth-ordering on shapes for statistical shape analysis.
- We develop an algorithm for the fast computation of band-depth for shapes as binary functions.
- We define statistical tests to detect potential global and local differences between shape populations.
- We provide the directionality of shape differences to augment local measurements.

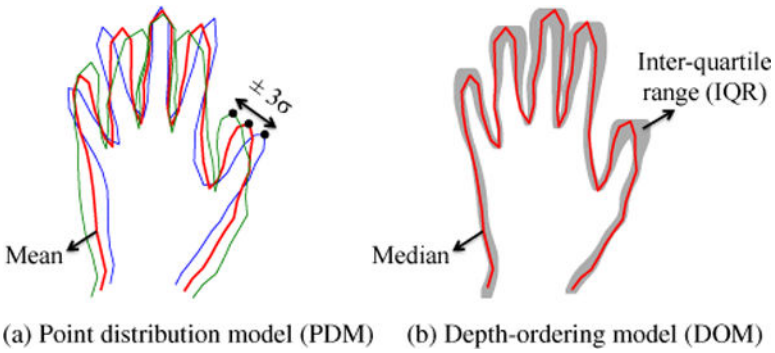

Figure 1.

Comparison between two types of models for capturing shape variations. The three shapes in PDM (a) correspond to the mean and varying shapes along the first mode at ±3 standard deviations. In DOM (b), the red shape is the median of the shape population, and the grey area is the region covered by 50% of the shapes at the top ranking list of the populations, similar to the inter-quartile range (IQR) for scalar values visualized as part of a box-plot.

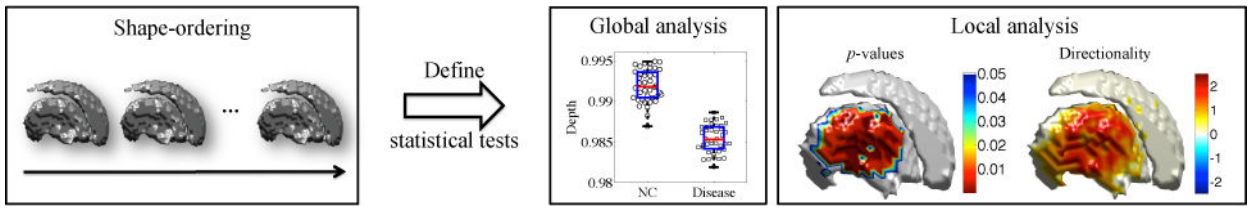

Figure 2.

Overview of the depth-ordering-based shape analysis. Based on the depth and the ordering of a shape population, statistical tests are defined to globally separate control and disease groups using global analysis with a scalar value (depth) for each shape. Statistical difference for global analysis results can be established through permutation testing. Equivalently, local shape differences can be detected using local analysis with a corresponding local permutation test, resulting in *p*-values on the surface of a shape to establish local shape differences between populations. The directionality of the shape differences (inflation versus deflation) can also be determined.

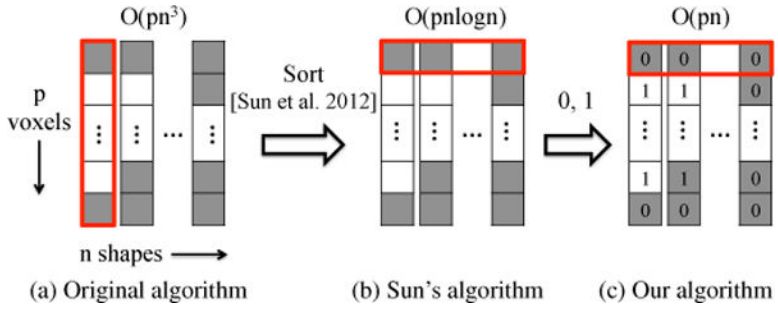

Figure 3.

Illustration of the computational complexity of band-depth calculation for all observations. The computational complexity of the original algorithm is cubic, and our algorithm reduces it to linear, with respect to the number of observations.

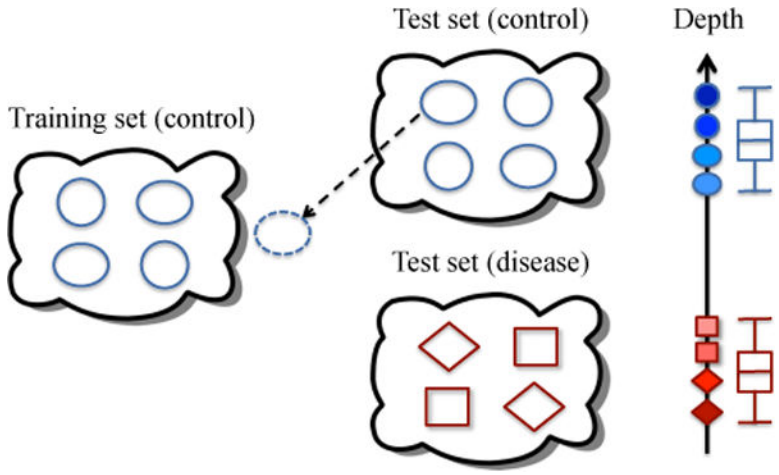

Figure 4.

Illustration of global shape analysis. A group of control subjects is chosen as a training set, and the band-depth of each test shape from control or disease groups, is computed with respect to the training set. The boxplots on the right show that in general the control group would have larger depths than the disease group if control subjects are selected as the reference/training population.

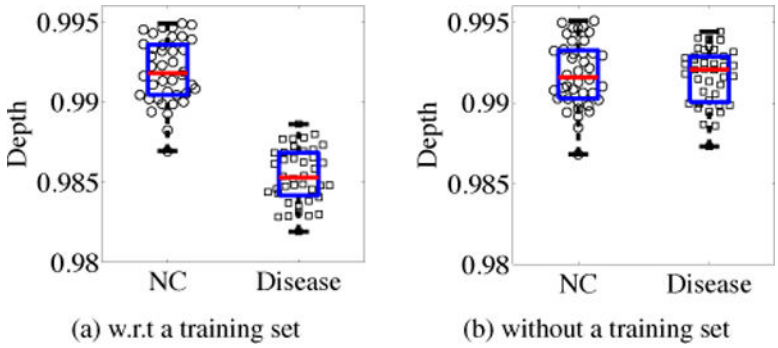

Figure 5.

Global shape analysis on synthetic data using band-depth computed with (a) and without (b) a training population. The training population allows detection of global shape differences.

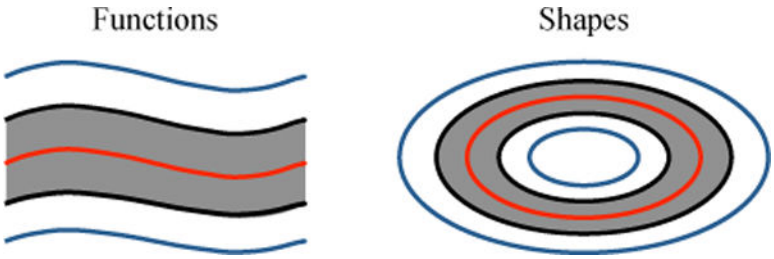

Figure 6.

Illustration of the median and α-central region. The median function / shape (red) is the most central one of a population, and the α-central region (grey) is the band delimited by the α proportion of the deepest functions / shapes. For example, the grey region is the 0.6-central region because it is built by three out of five functions / shapes.

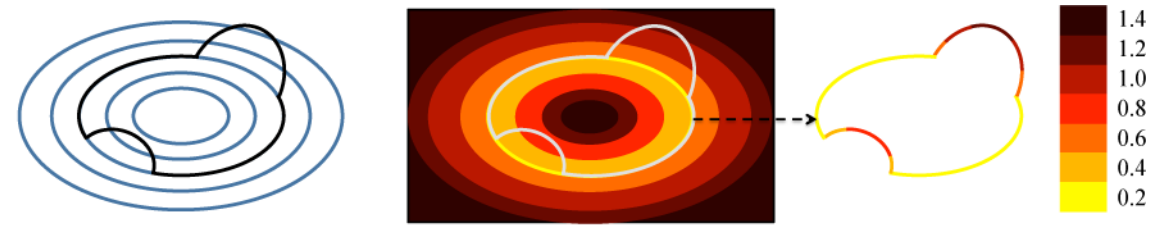

Figure 7.

Illustration of local shape analysis. Reference shape population (a) (blue contours) defines a centrality map (b) (light yellow to dark red corresponds to the most to the least central region of the reference population), which provides a local measure (c) (α-values) of how deeply a test shape (the non-ellipse shape) is buried with respect to the reference population. The dilated region is colored by the darkest red with α-values greater than 1.

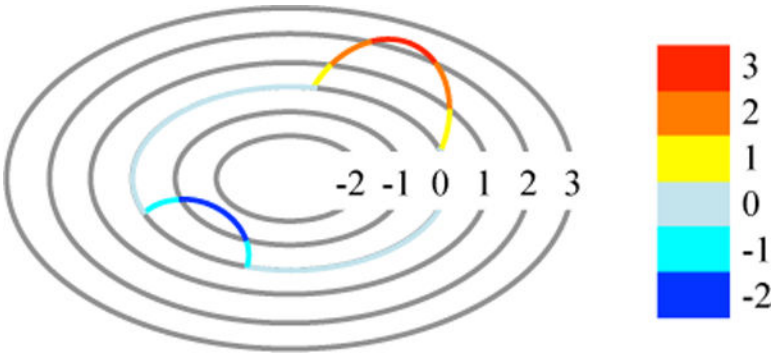

Figure 8.

Illustration of directionality for a template shape (the ellipse with a bump and an indented region) with respect to the median shape (the ellipse at the zero level-set) of a reference population.

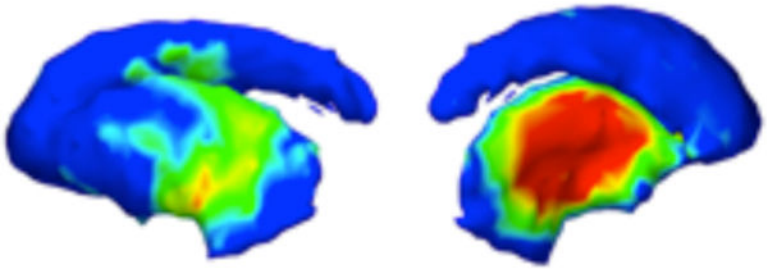

Figure 9.
Ground-truth of a sample shape (two views from the lateral and medial sides). Colormap on the shape indicates the location and magnitude of the artificial deformations compared to the undeformed shape. Red color corresponds to a large deformation distance.

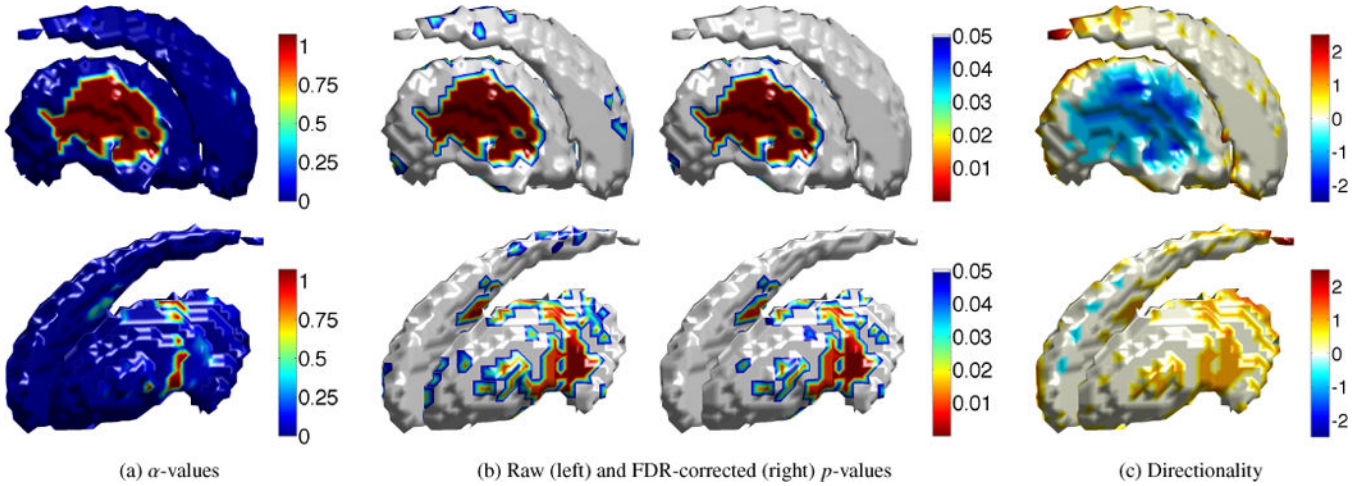

Figure 10.

Local shape analysis on synthetic striatum shown from two views, the medial (top) and the lateral (bottom) views, with the median of NC-Test as the template and the disease test group as the reference population. (a) The α-values on the template. (b) The corresponding raw *p*-values with 10000 permutations, and FDR corrected *p*-values. (c) The directionality of shape differences on the template with respect to the median of the reference group.

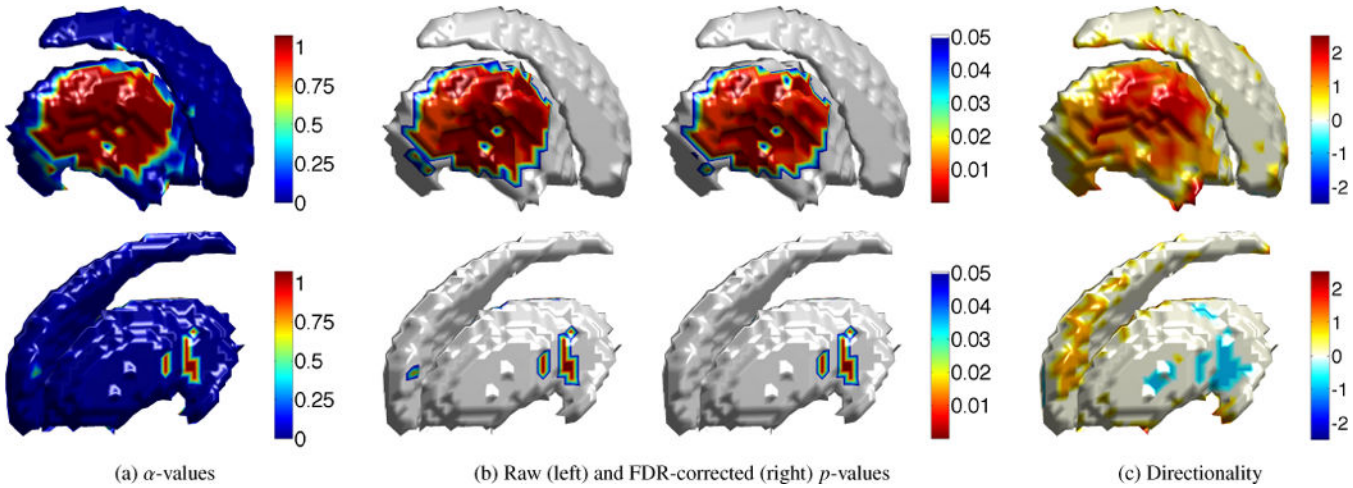

Figure 11.

Local shape analysis on synthetic striatum shown from two views, the medial (top) and the lateral (bottom) views, with the median of the disease test group as the template and the NC-Test group as the reference population. (a) The α-values on the template. (b) The corresponding raw *p*-values with 10000 permutations, and FDR corrected *p*-values. (c) The directionality of shape differences on the template with respect to the median of the reference group.

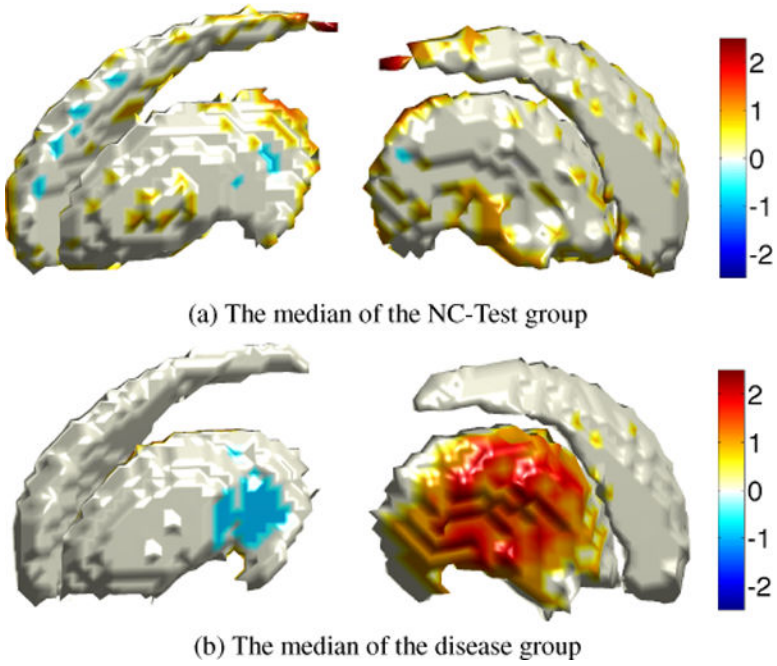

Figure 12.

The directionality of shape differences on the median test shapes with respect to the median shape of the NC-Train group shown from two views, lateral (left) and medial (right) sides.

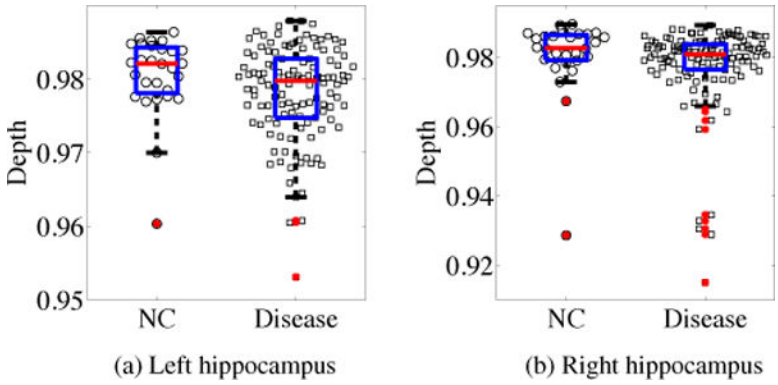

Figure 13.

Global analysis on both left and right hippocampi. The disease group indicates subjects with first-episode schizophrenia.

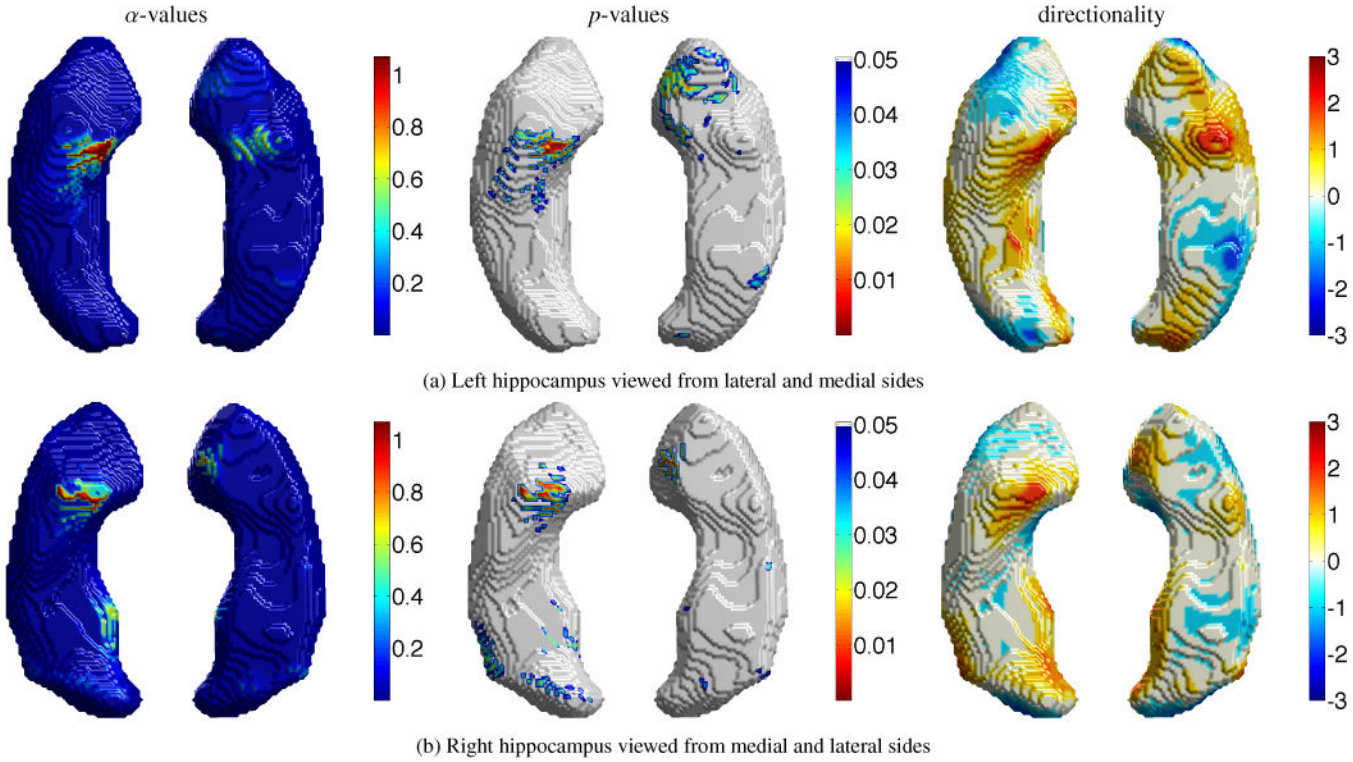

Figure 14.

Local analysis on the disease median of both left and right hippocampi with respect to the normal control test group.